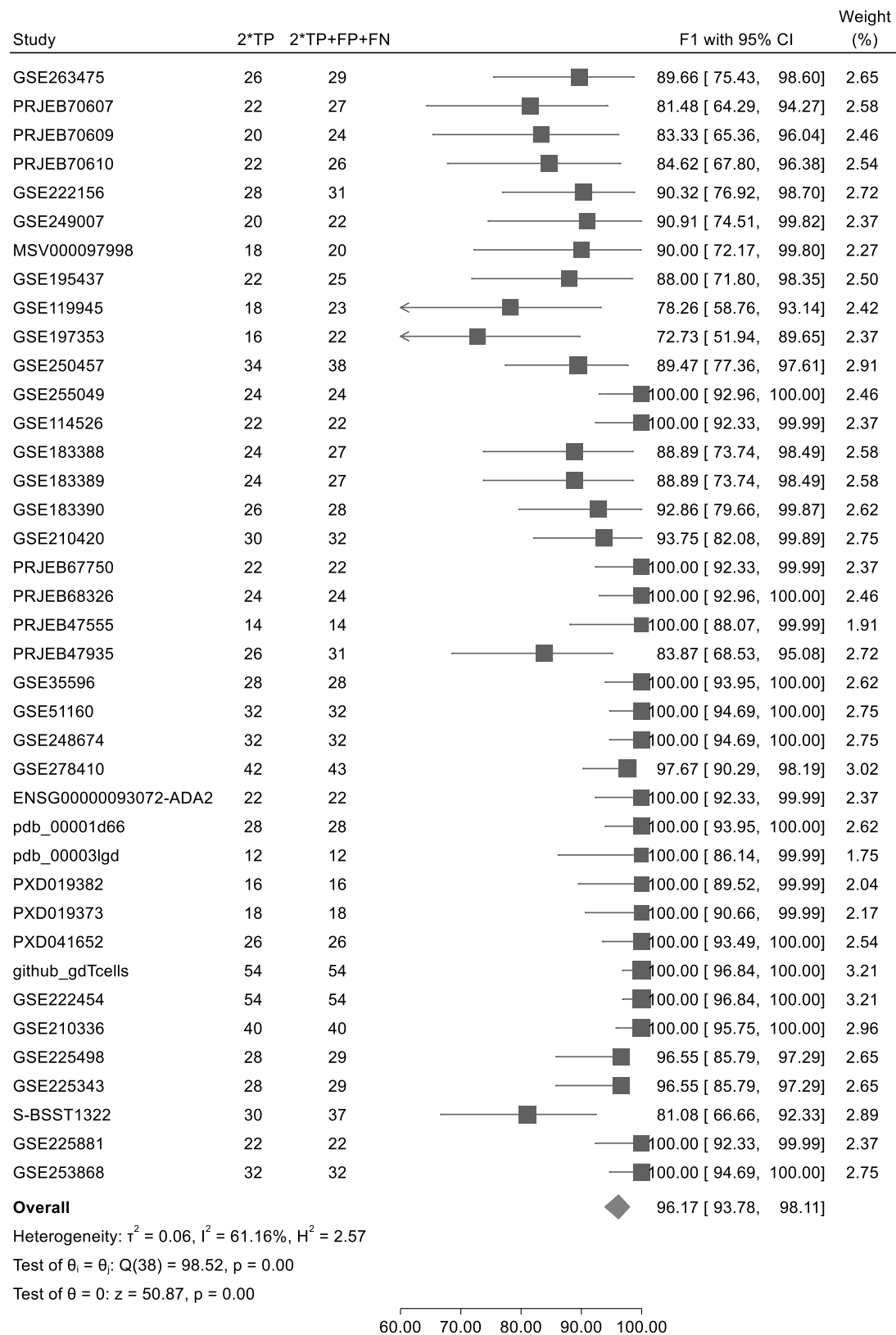


Supplemental Figure 1: F1 Score of the multi-step LLM workflow according to the human-in-the-loop verification of LLM suggestions



Random-effects REML model